*Ann. occup. Hyg.*, Vol. 47, No. 5, pp. 399–411, 2003
Published by Oxford University Press
DOI: 10.1093/annhyg/meg060

# Evaluation of Three Retrospective Exposure Assessment Methods

PATRICIA A. STEWART[1]*, PETER S. J. LEES[2], ADOLFO CORREA[2]†, PATRICK BREYSSE[2], MITCHELL GAIL[1] and BARRY I. GRAUBARD[1]

[1]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, 6120 Executive Boulevard, MSC 7240, Rockville, MD 20892-7240;* [2]*Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA*

Received 21 March 2002; in final form 7 March 2003

**Objective: To evaluate three methods for assessing retrospective exposures of acrylonitrile workers.**

**Methods: Three methods used to develop historical exposure estimates for a retrospective cohort mortality study of acrylonitrile workers were considered. The first method was deterministic, incorporating estimates of the impact of changes that took place in the workplace. The second method used the ratio of the mean of the measurements for three similar jobs to estimate a fourth similar job. The third method was based on the development of homogeneous exposure groups (HEG). Estimates of acrylonitrile exposure were developed using these three methods and compared with measurement means (observed means) across three categories of airborne exposure concentrations (<0.5, 0.5–0.99 and ≥1 p.p.m.) and three categories based on the number of measurements used to develop the estimate (<10, 10–29 and ≥30).**

**Results: The correlation between the estimates and the observed values was about 0.65 for all three methods. Estimates using the deterministic method tended to overestimate the observed means by 17%, but the number of estimates was not above or below the observed means more often than expected. There was no statistically significant relationship between the exposure estimates and the acrylonitrile concentration in the air or the number of measurements used to develop the estimates. The estimates averaged within 60% of the observed means when concentrations were above 0.5 p.p.m. and 25% regardless of the number of measurements on which the estimates were based. Estimates from the ratio method were randomly distributed above and below the observed means and averaged 70% above the observed means. The air concentration did not affect the performance of the method, although above 1 p.p.m. the estimates were within 40% of the observed means. The number of measurements comprising the estimates was related on a relative scale to the performance of the method. Exposure estimates using the HEG method were neither greater nor less than the observed means more often than what was expected. The method did better as concentration and the number of measurements increased. The estimates were within 60% of the means at >0.5 p.p.m. and for all measurement categories. Overall, there was no statistically significant difference between the estimates derived from the three estimation methods.**

**Conclusions: All methods performed reasonably well, but the deterministic and HEG methods appeared to develop estimates closer to the observed means for concentrations >0.5 p.p.m., regardless of the number of measurements.**

*Keywords:* acrylonitrile; dose reconstruction; epidemiological studies; epidemiology; exposure assessment; methods validation; reproducibility; validity

## INTRODUCTION

Most epidemiological studies of chronic disease lack the industrial hygiene measurement data necessary to estimate exposures for all jobs of interest or over the entire period of the study. Techniques are usually

*Author to whom correspondence should be addressed. Tel: +1-301-435-4714; fax: +1-301-402-1819; e-mail: stewartt@epndce.nci.nih.gov
†Present address: Center on Birth Defects and Developmental Disabilities, Atlanta, GA 30351, USA

developed to estimate exposures for these situations. Where measurements of concentrations are available for only a subset of the jobs, estimation of the exposures of unmeasured jobs may be done by comparing these jobs to one or more jobs with measurements. It is also common that industrial hygiene measurements are non-existent in the early years of interest. Estimation methods, therefore, are needed that can use current measurements to extrapolate exposure levels occurring in earlier time periods. Methods to estimate exposure levels have seldom been evaluated in terms of the accuracy of the exposure estimates derived (Stewart *et al.*, 1996a). Evaluation of such methods is needed to facilitate the selection of the most appropriate estimation method so as to minimize exposure misclassification in epidemiological studies.

In a recent epidemiological study on the mortality of acrylonitrile workers, three exposure estimation techniques were developed to estimate past occupational exposures to airborne acrylonitrile (Stewart *et al.*, 1996b). In that study, these methods were evaluated by comparing the derived exposure estimates with existing airborne acrylonitrile exposure measurement data by work site and by type of industrial process. Such an analysis is useful for interpreting individual study results, but provides little information for other studies. This paper compares the performances of the three estimation methods used in that study and explores in greater depth how each method performed. This analysis was undertaken independently of the exposure assessment effort of the epidemiological study, i.e. the estimated results from this validation study were not used in the epidemiological study.

## MATERIALS AND METHODS

A basic description of the three exposure estimation methods evaluated here has been provided elsewhere (Stewart *et al.*, 1996b). These methods were modified slightly for the analysis presented in this paper. Briefly, the years over which measurements existed were divided into time periods for each job. During each time period no changes in the workplace were known to have occurred that might have affected the exposures of a job, so that the exposures were thought to have been relatively constant. The strategy for the analysis described in this article was for an industrial hygienist to develop from a subset of the measurement data ('estimation data set') exposure estimates similar to those in the epidemiological study without knowledge of the measurement data ('referent data set'). The unused measurement data could then serve as the referent standard, or the 'truth', for this evaluation. The measurements used in this validation study had been collected by company personnel from 1977 to 1987 in the breathing zone of the employees for at least 6 h. A variety of sampling and analytic methods were used at the companies, but they were primarily measured by charcoal tubes with gas chromatographic analysis. The purpose of taking the measurements varied. More detail on the measurements is provided elsewhere (Zey *et al.*, 2002).

### *Deterministic method*

A deterministic method was used to estimate exposure levels in jobs that had been measured in some time periods but not in others, by taking into account the effect various changes in the workplace had on exposure concentrations over time. For the present analysis, four types of exposure determinants were identified that, if changed, could have resulted in changes in exposure concentrations: operating equipment, engineering controls, work practices and production rates. [In the earlier evaluation, production rates were excluded from the model (Stewart *et al.*, 1996b).] When one of these determinants changed, a new estimate was developed for all affected job/department/plant combinations (hereafter called job) existing in the relevant time period.

Typically, estimation of historical exposures is performed by starting with a mean of the measurement data in the most recent time period and working backwards through time. This procedure was followed here. The calculation for developing the new estimate worked backwards cumulatively using the formula:

$$E_{t-1} = E_t/(1 - \Sigma P_t + \Sigma P_t/R_t)$$

where $E$ is the exposure level at time $t$ for $t = 0, -1, -2$, etc. $E_t$ is the mean of the measurement data from the estimation data set for a job at time 0, which corresponds to the current or most recent time period with exposure measurements. We suppose that a portion $P_t$ of the exposure level is subject to process modification and that $R_t$ is the ratio of the exposure level in that portion at time $t - 1$ to the exposure at time $t$. Then $E_{-1} = E_0/(1 - \Sigma P_0 + \Sigma P_0/R_0)$, $E_{-2} = E_0/(1 - \Sigma P_{0,1} + \Sigma P_{0,1}/R_{0,1})$, and so forth. For each change, a weighting factor ($F$) was computed as $PR$, where $P$ and $R$ are as defined above. The weighting factor for a particular job/time period reflected both the change that occurred for the particular time period being estimated and, in addition, all of the more recent changes, so that as the exposure levels increased backwards through time, the magnitude of the factor increased. Generally, $P$ and $R$ were estimated from knowledge of the job tasks and work environment, published reports on similar operations and personal experience. Few were based on actual measurement data.

To evaluate the estimation methods, means of measurements (hereafter called the observed means) from the referent data set were calculated for each job/time period. Other than for the initial baseline estimate, each subsequent estimate was compared with the corresponding observed mean for the same job and time period, the latter serving as the referent standard. When no measurement data existed for an estimate, the estimate was deleted from the comparison analysis.

For example, assume in 1978 the baseline estimate ($E_0$) for the reactor operator was 1.64 p.p.m. In 1977 OSHA issued an emergency temporary standard that caused the employer to be more aware of acrylonitrile exposures. Assume this awareness reduced the 1977 exposures by 70% ($R_0 = 3.5$) and affected 10% of the reactor operator's exposure ($P_0 = 0.1$). Assume for simplicity that there was no change in production rates. The estimated concentration of $E_{-1} = 1.64/(1 - 0.1 + 0.1/3.5) = 1.77$ p.p.m., which was compared with the observed mean exposure for that job/time period (3.19 p.p.m.). The value for the weighting factor for this change was calculated as $0.1 \times 3.5 = 0.35$.

*Ratio method*

The second estimation method, the ratio method, was developed to estimate exposure levels of jobs that had never been measured. It is based on the assumption that jobs performing similar tasks in similar environments in different work sites are likely to have similar exposure concentrations relative to other jobs in the same operation (Stewart *et al.*, 1996b). Expressed mathematically, this assumption is:

$$\mathrm{job}_{ij}/\mathrm{job}_{i'j} = \mathrm{job}_{ij'}/\mathrm{job}_{i'j'}$$

where the first subscript identifies the *i*th job, the second indicates the *j*th work site and the prime designates a second job or work site. All four jobs had to have been monitored at least once. The means for the three jobs, derived from the estimation data set, were used to calculate the exposure level for the fourth job and compared with the mean of that job from the referent data set. For example, assume in 1977 the same reactor operator described for the deterministic method worked with a salt purifier (mean in 1977 2.24 p.p.m.) in the same department and another similar reactor operator (mean 0.60 p.p.m.) worked with a salt purifier (mean 0.82 p.p.m.) under similar conditions with similar controls at a similar work site. The last three jobs were used to estimate the 1977 reactor operator's exposure at the first work site as $2.24 \times 0.60/0.82 = 1.64$ p.p.m., which was compared with the observed mean for that job (3.19 p.p.m.).

Four estimates were developed from each unique set of four jobs, one for each job. When there were more than four jobs within the same plant or in other plants that met the inclusion criteria, all possible combinations of jobs were evaluated, developing four estimates for each combination.

*Homogeneous exposure group method*

The third estimation method used in the epidemiological study was called the homogeneous exposure group (HEG) method, developed to estimate exposure levels for jobs that had never been measured. It was based on the concept that workers working in the same area for similar amounts of time are likely to have similar exposure levels. In this study, HEGs were developed when there were at least three jobs with expected similar exposures and with at least one measurement each. The jobs in the HEG had to have been carried out in the same physical area of the plant for approximately the same number of hours per day and were likely to have been affected by the same sources of emissions and controls. Ninety HEGs were developed across all of the plants in the study. For each HEG, after excluding the referent measurements of one job, a mean was calculated of the individual measurements from the estimation data set on the remaining jobs in the HEG. This mean became the estimate for the exposure level of the excluded job and was compared with the observed mean from the referent data set for this job. Within each HEG, estimates were developed for all possible jobs. For example, in 1977 the reactor operator (mean 3.19 p.p.m., $n = 21$), the salt handler (mean 2.83 p.p.m., $n = 30$) and the salt purifier (mean 2.24 p.p.m., $n = 24$) all worked in the same location at a single plant. The means of the measurements for the salt handler and the salt purifier were used to estimate the exposure level of the reactor operator (2.57 p.p.m., $n = 54$); those of the salt purifier and the reactor operator were used to estimate the exposure level of salt handler (2.68 p.p.m., $n = 45$); those of the reactor operator and the salt handler were used as an estimate for the salt purifier (2.98 p.p.m., $n = 51$). Each of these estimates was compared with its corresponding observed mean.

*Statistical methods*

In the epidemiological study for which these data were collected, the disease outcome of interest was cancer. When a linear exposure–response model is thought to be appropriate to the disease being studied (such as for acrylonitrile and cancer), the arithmetic mean results in less bias than the geometric mean (Seixas *et al.*, 1988). For this reason, we describe arithmetic means in this paper.

Included in the statistical analysis are the estimates for only those jobs that were developed from all three methods (Appendix), to allow comparison of the performance of the methods and determination of the magnitude of the differences between the estimation methods. Several statistics were calculated. (i) The Spearman correlation coefficient was calculated to determine the association between the individual estimates and their corresponding observed means. This statistic is important because correct ranking of jobs is needed for an appropriate interpretation of exposure–response relationships in epidemiological studies. (ii) To determine if the exposure estimates were lower or higher than the measured concentrations more often than was expected due to chance, the Wilcoxon ranked sign test was used (Snedecor and Cochran, 1973).

In addition, three agreement statistics were calculated. (i) Estimates may be highly correlated and the direction of the estimates may be random and yet the estimates may not be reflective of the true exposure levels. The difference between the observed mean and its corresponding estimate was calculated for each job/department/plant/time period. A mean of these differences was used to describe the average difference between the two sets of values. In the text the percent of the difference is also used to describe the method's performance. The percentage was calculated by subtracting the difference between the observed mean and the estimate and dividing by the observed mean. (ii) Because the average difference can be deceptive, in that the positive and negative differences can cancel each other out, the average absolute difference was calculated to estimate the magnitude of the differences between the observed mean and its corresponding estimate, without regard to the direction of the difference. (iii) The average relative absolute difference, i.e. the average of the absolute differences, each difference divided by its corresponding observed mean, is a measure of relative error. The closer to 0.00 this statistic is, the less misclassification there is. We assumed that any relative average difference <2.00 is good performance, as this implies that the absolute differences were equal to or less than 2-fold the means. Large average absolute differences and large average relative absolute differences indicate that some of the differences between the observed means and the estimates are large, suggesting that the precision of the estimates is poor. The relative statistic is also useful for comparison across categories with widely differing exposure levels.

Each agreement statistic is presented together with its standard error. For the deterministic method, a simple standard error of the mean is presented. For the other two methods, the standard error was calculated using a bootstrap method (Efron and Gong, 1983) to account for the lack of independence of the observations contributing to the estimates.

To calculate the standard error for the ratio method, all jobs were assigned to a 'net'. As indicated above, when there were more than four jobs within the same plant or in other plants that met the inclusion criteria, all possible combinations of jobs were evaluated. A net comprised a unique set of jobs that were associated with each other (having been used in at least one ratio calculation), but did not include any job used in a different net. This resulted in 16 independent nets. The three agreement statistics were calculated for each job within the net and a mean of each of the agreement statistics was calculated for the net. A random selection with replacement of 16 ('bootstrap') samples was made of the 16 nets. This selection of 16 samples comprised a bootstrap sample and was repeated 100 times. Means and standard errors ($SE_b$) of the agreement statistics were calculated from the 100 bootstrap samples. Thus, although there was correlation among the jobs in each net, the 'bootstrap' standard error estimates allowed for such correlations by resampling independent nets.

To calculate the standard error of each of the mean statistics for the HEG method, the agreement statistics were calculated for all of the jobs within the HEG to derive an HEG set of agreement statistics. A bootstrap sample of 90 HEGs was made with replacement from the 90 HEGs. This procedure of taking 90 random samples was repeated 100 times. The means of these 100 agreement statistics were calculated together with their bootstrap estimate of standard error, $SE_b$.

The performance of an estimation method may be dependent on variables out of the control of the industrial hygienist making the estimates. To determine if the concentration of the observed means influenced the performance of the methods, the statistics described above were calculated separately for three exposure concentration categories (<0.5, 0.5–0.99 and >0.99 p.p.m.). The effect of the number of acrylonitrile measurements used to develop the estimates (<10, 10–29 and ≥30 measurements) was also evaluated. For the deterministic method, evaluation was also made as to the magnitude of the weighting factor ($F$) used to modify the estimates and the number of generations over which the original estimate had been modified. Only two generations (the first and third) were evaluated because of the limited number of comparisons. The paired $t$-test was used to compare the mean of the differences between generations. These two variables (i.e. weighting factors and generations) included all comparisons from the deterministic method ($n = 340$ and 44, respectively), not just those that had corresponding ratio or HEG estimates, because these two variables were not applicable to the other two estimation methods.

To compare the methods' performances, we used the paired $t$-test, which tested for a difference between each pair of corresponding estimates. To apply this test to two methods, we first calculated the difference for each pair of estimates from the two methods, which produced three sets of 59 differences (i.e. from 59 differences between the deterministic and ratio methods, 59 differences for the ratio and HEG methods and 59 differences for the deterministic and HEG methods). Because the estimates from the ratio method were related within nets but were not correlated across the nets, we averaged the differences within each net for each pair of methods to produce 16 mean differences. These averages were then averaged for the paired $t$-test.

All statistical tests were two-sided. The values used for measurement results below the limit of detection (LOD) of the method were calculated by dividing the LOD by $\sqrt{2}$ (Hornung and Reed, 1990). About 20% of the measurements were below the LOD.

## RESULTS

### *The deterministic method*

There was moderate correlation ($r = 0.63$, $P < 0.001$) between the observed means and the exposure estimates derived from the deterministic method. The overall average difference between the observed means and the estimates was –0.18 p.p.m. [–17% (= –0.18/1.08 p.p.m.)] (Table 1) and the average relative absolute difference (1.01) indicates that there were relatively few large differences between the observed means and the estimates. The number of estimates that exceeded the observed means was greater than expected ($P < 0.04$).

These overestimates occurred at lower concentrations (<1 p.p.m.) (Table 2). There was no statistical difference in the direction of the estimates at or above 1 p.p.m. There was also no statistically significant correlation between the concentration of the observed means and the magnitude of the differences ($r = 0.22$, $P = 0.09$) or the magnitude of the relative absolute differences ($r = -0.19$, $P = 0.16$) (not shown). Overall, the estimates were within 60% of the observed means above concentrations of 0.5 p.p.m. (–0.36/0.70 and 0.08/1.93 p.p.m. for the 0.50–0.99 and ≥1 p.p.m. categories, respectively) and the relatively low average relative absolute differences indicate that there were few large differences in these two categories. Thus, the largest differences appear to have occurred in the lowest exposure category, as is indicated by the larger average relative absolute difference of 1.88 (compared with 0.67 and 0.56 for the other two categories). Correlations within each concentration category were 0.30, 0.25 and –0.01 for the <0.5, 0.5–0.99 and ≥1 p.p.m. categories, respectively.

No pattern was evident between the number of measurements on which the estimates were based and the differences and relative absolute differences ($r = -0.20$, $P = 0.12$ and $r = -0.03$, $P = 0.81$, respectively) (not shown). However, the method generally performed well regardless of the number of measurements. For all categories of measurements (<10, 10–29 and ≥30) the estimates averaged within 25% of the observed means. The method performed somewhat better on a relative scale when the number of measurements equaled 10 or more. Correlations within each category were –0.06, 0.95 and 0.21 for the <10, 10–29 and ≥30 categories, respectively.

When the data were examined by the magnitude of the weighting factor, the estimates were more often underestimates for the smaller weights ($F \leq 3.00$) and overestimates for the larger weights ($F > 3.00$). The correlation coefficient showed a negative association between the differences and the factors, i.e. as the value of the factor increased, the estimates increasingly tended to overestimate the observed ($r$ for the differences = –0.49, $P < 0.001$ and for the relative absolute differences, $r = 0.51$, $P = <0.001$) (not shown). The average and the average relative absolute differences between the observed means and the estimates also increased at higher weighting factors, particularly when the factor exceeded six. This is confirmed by the generation analysis. The average difference increased from essentially 0.0 p.p.m. at generation 1 to 1.28 p.p.m. at generation 3. On average, the relative absolute estimates in the third generation were 2.5-fold larger than the observed means. The difference between the generations, however, was not statistically significant ($t = -1.14$, $P = 0.27$).

### *The ratio method*

The correlation of the observed means and the estimates developed from the ratio method was 0.64 ($P < 0.001$). The overall average difference between the observed and the estimates was –0.71 p.p.m. (66% of the observed means) (Table 1) and there were some

Table 1. Statistics (mean ± SE) for the differences between the means of the exposure measurements and the estimates derived from the three estimation methods

| Estimation method (observed mean 1.08 p.p.m.) | $n$ | Average difference | | Average absolute difference | | Average relative absolute difference | |
|---|---|---|---|---|---|---|---|
| | | Difference (p.p.m.) | (SE)[a] | Difference (p.p.m.) | (SE) | Difference | (SE) |
| Deterministic | 59 | –0.18[b] | (0.16) | 0.73 | (0.13) | 1.01 | (0.21) |
| Ratio | 59 | –0.71 | (0.05) | 1.15 | (0.05) | 1.58 | (0.07) |
| HEG | 59 | –0.03[c] | (0.01) | 0.60 | (0.02) | 1.08 | (0.03) |

[a]The standard error for the deterministic method was the standard error of the mean. The standard error for the ratio and HEG methods were bootstrap standard errors.
[b]The results of the Wilcoxon ranked sign test were statistically significant ($P < 0.05$).
[c]This value is uninformative (see Results).

Table 2. Statistics (mean ± SE) for the differences between the means of the exposure measurements and the estimates derived from the deterministic method

| Type of evaluation (observed mean)[a] | $n$ | Average difference | | Average absolute difference | | Average relative absolute difference | |
|---|---|---|---|---|---|---|---|
| | | Difference (p.p.m.) | (SE) | Difference (p.p.m.) | (SE) | Difference | (SE) |
| *Concentration*[b] | | | | | | | |
| <0.50 p.p.m. (0.26 p.p.m.) | 19 | –0.39[c] | (0.16) | 0.44 | (0.16) | 1.88 | (0.58) |
| 0.50–0.99 p.p.m. (0.70 p.p.m.) | 15 | –0.36[d] | (0.19) | 0.50 | (0.17) | 0.67 | (0.21) |
| ≥1 p.p.m. (1.93 p.p.m.) | 25 | 0.08 | (0.34) | 1.08 | (0.26) | 0.56 | (0.12) |
| *No. of measurements* | | | | | | | |
| <10 (0.59 p.p.m.) | 13 | 0.15 | (0.26) | 0.54 | (0.21) | 1.46 | (0.58) |
| 10–29 (0.51 ppm) | 7 | –0.08 | (0.06) | 0.12 | (0.04) | 0.52 | (0.28) |
| ≥30 (1.34 p.p.m.) | 39 | –0.31 | (0.23) | 0.90 | (0.18) | 0.95 | (0.25) |
| *Weighting factor*[e] | | | | | | | |
| <1 (0.47 p.p.m.) | 60 | 0.19[c] | (0.07) | 0.38 | (0.07) | 0.76 | (0.06) |
| 1.01–3.00 (1.25 p.p.m.) | 216 | 0.42 | (0.18) | 1.01 | (0.17) | 1.83 | (0.32) |
| 3.01–6.00 (1.21 p.p.m.) | 42 | –0.83[f] | (0.10) | 1.47 | (0.07) | 2.64 | (0.17) |
| >6 (0.42 p.p.m.) | 22 | –1.36[f] | (0.06) | 1.40 | (0.06) | 5.07 | (0.30) |
| *Generation*[g] | | | | | | | |
| 1 (0.82 p.p.m.) | 22 | 0.03 | (0.06) | 0.58 | (0.05) | 0.73 | (0.05) |
| 3 (2.87 p.p.m.) | 22 | 1.28 | (0.31) | 2.94 | (0.27) | 2.50 | (0.24) |

[a]For the analyses of the concentration and the number of measurements, $n = 59$ to allow comparison with Tables 3 and 4. For the weighting factor and the generation analyses, the number of comparisons included all possible comparisons because these analyses are not relevant to the other two methods.
[b]From the mean of the measurements.
[c]The results of the Wilcoxon ranked sign test, $P < 0.01$.
[d]The results of the Wilcoxon ranked sign test, $P < 0.05$.
[e]Incorporating the portion of the exposure affected and the magnitude of the effect.
[f]The results of the Wilcoxon ranked sign test, $P < 0.001$.
[g]The number of estimates developed back through time.
Significance was determined by a 2-sided $t$-test.

Table 3. Statistics (means ± SE) for the differences between the means of the exposure measurements and the estimates derived from the ratio method

| Type of evaluation (observed mean) | $n$ | Average difference | | Average absolute difference | | Average relative absolute difference | |
|---|---|---|---|---|---|---|---|
| | | Difference (p.p.m.) | $(SE_b)$[a] | Difference (p.p.m.) | $(SE_b)$ | Difference | $(SE_b)$ |
| *Concentration*[b] | | | | | | | |
| <0.5 p.p.m. (0.26 p.p.m.) | 19 | –0.66 | (0.05) | 0.78 | (0.05) | 2.66 | (0.14) |
| 0.5–0.99 p.p.m. (0.70 p.p.m.) | 15 | –0.66[c] | (0.04) | 0.71 | (0.04) | 0.95 | (0.05) |
| ≥1 p.p.m. (1.93 p.p.m.) | 25 | –0.78 | (0.10) | 1.68 | (0.10) | 1.24 | (0.09) |
| *No. of measurements* | | | | | | | |
| <10 (0.40 p.p.m.) | 7 | 0.12 | (<0.01) | 0.23 | (0.01) | 1.65 | (0.11) |
| 10–29 (0.59 p.p.m.) | 13 | –3.43 | (0.30) | 3.79 | (0.29) | 5.32 | (0.38) |
| ≥30 (1.36 p.p.m.) | 39 | –0.09 | (0.01) | 0.53 | (0.01) | 0.44 | (0.01) |

[a]$SE_b$, standard error derived from the bootstrap method.
[b]From the mean of the measurements.
[c]The results of the Wilcoxon ranked sign test, $P < 0.01$.

large differences between the observed means and the estimates. Overall, the direction of the differences appeared to be random ($P = 0.24$). In the middle concentration category, however, the number of the estimates exceeding the observed means was statistically significant (Table 3). There was no statistically significant correlation between the observed means and either the differences or relative absolute differences ($r = 0.18$, $P = 0.17$ and $r = -0.25$, $P = 0.05$, respectively) (not shown), although the latter

Table 4. Statistics for the differences between the means of the exposure measurements and the estimates derived from the homogeneous exposure group method

| Type of evaluation (observed mean) | $n$ | Average difference | | Average absolute difference | | Average relative absolute difference | |
|---|---|---|---|---|---|---|---|
| | | Difference (p.p.m.) | $(SE_b)$[a] | Difference (p.p.m.) | $(SE_b)$ | Difference | $(SE_b)$ |
| *Concentration*[b] | | | | | | | |
| <0.5 p.p.m. (0.26 p.p.m.) | 19 | –0.25[c] | (0.01) | 0.33 | (0.01) | 2.17 | (0.09) |
| 0.5–0.99 p.p.m. (0.70 p.p.m.) | 15 | –0.40[d] | (0.02) | 0.44 | (0.02) | 0.60 | (0.02) |
| ≥1 p.p.m. (1.93 p.p.m.) | 25 | 0.33 | (0.01) | 0.95 | (0.04) | 0.51 | (0.02) |
| *No. of measurements* | | | | | | | |
| <10 (0.51 p.p.m.) | 15 | –0.16 | (0.03) | 0.57 | (0.02) | 2.43 | (0.11) |
| 10–29 (0.76 p.p.m.) | 11 | –0.03 | (0.01) | 0.36 | (0.01) | 0.60 | (0.02) |
| ≥30 (1.4 p.p.m.) | 33 | –0.02 | (0.01) | 0.69 | (0.03) | 0.61 | (0.02) |

[a]$SE_b$, standard error derived from the bootstrap method.
[b]From the mean of the measurements.
[c]The results of the Wilcoxon ranked sign test, $P < 0.05$.
[d]The results of the Wilcoxon ranked sign test, $P < 0.01$.

relationship was marginally statistically significant. Only in the highest exposure category (≥1 p.p.m.) did the estimates average within 40% of the observed mean. The absolute differences from the ratio method on a relative scale were about 2.5 times the observed means in the lowest exposure category on average, but came closer to estimating the observed means in the other two categories (95 and 124%). Correlations within each concentration category were 0.32, 0.31 and 0.11 for the <0.5, 0.5–0.99 and ≥1 p.p.m. categories, respectively.

There was no trend in differences associated with the number of measurements and the differences ($r = 0.08$, $P = 0.57$), but there was a negative statistically significant correlation between the number of measurements and the relative absolute differences ($r = -0.50$, $P < 0.001$), indicating that as the number of measurements increased, the relative absolute differences decreased. For the middle category the precision of the estimates was poor. Correlations within each category were 0.73, –0.15 and –0.78 for the <10, 10–29 and ≥30 categories, respectively.

*The homogeneous exposure group method*

The Spearman correlation coefficient between the observed means and the estimates was 0.66 ($P < 0.001$) for the HEG method and the average difference was 0.03 p.p.m. (Table 1). A value close to 0 was expected, however, because of the procedure used to calculate the estimates. For example, if the three jobs described earlier comprised the HEG, the differences between the means and the estimates are: for the reactor operator, 2.83 – 2.72 = 0.11 p.p.m; for the salt purifier, 2.24 – 3.01 = –0.77; for the salt handler, 3.19 – 2.54 = 0.65. When averaged the overall difference is about 0 (0.01 p.p.m.). This phenomenon occurs for all the HEGs and therefore this calculation is uninformative. There were few large differences between the observed means and the estimates (average relative absolute difference 1.08), and the direction of the estimates was randomly distributed ($P = 0.35$).

The magnitude of the observed means was moderately associated with the differences ($r = 0.47$, $P < 0.001$) and inversely associated with the relative absolute differences ($r = -0.33$, $P < 0.01$) (not shown). The number of estimates that exceeded the observed means was greater than expected in the lowest two exposure categories (Table 4). The estimates overestimated the observed by 0.25 p.p.m. (96%) for the lowest exposure category. At or above 0.5 p.p.m. the estimates were within 60% of the observed mean and at or above 1 p.p.m. they were randomly distributed around the means. Correlations within each concentration category were 0.35, 0.47 and –0.04 for the <0.5, 0.5–0.99 and ≥1 p.p.m. categories, respectively. There was no correlation between the number of measurements comprising the observed means and the differences ($r = -0.02$, $P = 0.88$), but the correlation was negative with the magnitude of the relative absolute differences ($r = -0.31$, $P = 0.02$). For all measurement categories the estimates averaged within about 30% of the observed mean and, on a relative scale, the average absolute difference fell from 243% for fewer than 10 measurements to 60 and 61% when there were ≥10 and ≥30 measurements, respectively. Correlations within each category were 0.20, 0.21 and 0.22 for the <10, 10–29 and ≥30 categories, respectively.

Overall, the estimates derived from the three estimation methods were not statistically different (Table 5), based on the paired $t$-test. The deterministic and HEG performed similarly, but there were larger differences between the ratio method and the other two estimation methods.

Table 5. Statistics for the differences between the means of the measurements and the estimates derived from the three methods using the same jobs for comparisons[a]

| Method ($n$ = 16) | Average difference | | Average absolute difference | | Average relative absolute difference | |
|---|---|---|---|---|---|---|
| | Difference (p.p.m.) | (SE)[b] | Difference (p.p.m.) | (SE) | Difference (p.p.m.) | (SE) |
| DET–HEG | −0.10 | (0.21) | 0.06 | (0.21) | −0.52 | (0.52) |
| DET–Ratio | 1.04 | (1.08) | −0.87 | (1.09) | −1.20 | (1.48) |
| Ratio–HEG | −1.14 | (1.02) | 0.94 | (1.07) | 0.68 | (1.47) |

[a]No difference (average, average absolute or average relative absolute difference) was statistically significantly different from 0. Calculated from the average of the paired (i.e. the two methods being compared) differences in each net.
[b]See Table 1 for the definition.

## DISCUSSION

Industrial hygienists typically must develop exposure estimates for retrospective epidemiological studies in the absence of measurement data. The credibility of the study often rests on the exposure assessment, but validation of those estimates has largely remained unexplored. The purpose of this paper was to evaluate three exposure estimation methods to determine how well they developed estimates compared with measurement data.

Correlations between the estimates and the measurement results for all three methods were about 0.65 and comparison of the methods found little difference between the estimates developed by the deterministic method and the HEG method. In contrast, there were larger differences with the ratio method, although the differences were not statistically significantly different from those of the other two methods. There was little power, however, to distinguish among the methods. This is, in part, because all three of the methods relied heavily on the same information obtained from interviews of waged and salaried employees years after the occurrence of the exposures. There was, therefore, likely to be some correlation among the estimates derived from the same data collection effort. Because we were unable to measure or estimate this correlation, we selected a conservative approach of estimating the standard error. In spite of the non-significant findings, given the choice we would recommend the HEG or deterministic method over the ratio method to extrapolate exposures to other jobs and time periods, respectively, because they performed similarly and better than the ratio method in most of the statistics calculated.

It is useful to know how well the methods worked under different circumstances, because industrial hygienists do not have the luxury of selecting the conditions under which they develop their estimates. We examined the comparison of the estimates and the measurement results at different levels of exposure and with different numbers of measurements to determine if the methods worked better under some circumstances than others.

### *Deterministic method*

The deterministic model relied on the assumptions that the most important determinants of changes in airborne exposures are production rates, the use of engineering controls and the occurrence of other process and work practice changes and that the impact of these variables on exposures could be accurately estimated. The assumption that production rates affect exposures was based on work by Esmen (1979). In the operations in this study, production volumes generally increased over time. Thus, if there were no relation between the production output and exposures or if the relationship is weaker than indicated by Esmen, there will be a systematic upward bias, i.e. the estimates developed for historical times will be higher than the actual concentrations were. If the impact of production rate is stronger than the values used in the study, the historical estimates will be less than the actual exposure levels were. The production rates were annual figures provided by the company and should be correct. The other data (engineering controls and other changes) were estimated from knowledge of the workplace and are likely to be more prone to error, which is assumed to be random.

Overall, the deterministic method performed well both for the average difference and the average relative absolute difference, although the method tended to overestimate the observed means. In addition, the correlation between the observed means and the estimates was moderate. There was no statistically significant effect of air concentration on performance of the method, although there was some indication that the method performed better at concentrations exceeding 0.5 p.p.m., where the estimates averaged within 60% of the observed means. There was also no significant relationship with the differences and the number of measurements used to develop the estimates, with the estimates averaging within about 25% of the observed means for all measurement categories. There was, however, some indication that the method performed better with more measurements (≥10). In contrast, as the weighting factors increased in magnitude and the number of generations increased, the method performed worse. This is not unexpected because as the weighting factors and number of generations increased, the uncertainty increased.

Of the 59 estimates developed using the deterministic method, the difference between the observed means and the estimates of 13 comparisons was >1 p.p.m. Of these, 70% had exposure means different from what might have been expected, based on either a comparison with the means from other years for the same job or of other jobs in the same year and location. For example, in 1979 the assistant polymer operator in plant 1 had a mean exposure of 7.36 p.p.m. ($n = 24$). One time period prior to and two time periods after that the mean exposure was 1.22–1.58 p.p.m. ($n = 9–25$). Other nearby jobs had mean exposures of 1.05–1.30 p.p.m. ($n = 24–28$) in 1979. Thus it is possible that the mean of the assistant polymer operator is unusually high. The deterministic method, therefore, appears to perform poorly when exposure concentrations do not follow expected patterns.

*Ratio method*

The ratio model assumed that the relative exposure concentration between related jobs is the same across all workplaces with similar conditions. This assumption appears to be reasonable in light of work by Eisen *et al.* (1984), in which the exposures of six jobs in seven granite sheds were compared. Although exact exposure concentrations were not provided in that study, a graph was presented that suggested that the exposures of the six jobs were related in a multiplicative fashion across the sheds, even when the exposure levels varied in concentration across the sheds.

Overall, the method developed estimates that averaged within 70% of the observed means and that were randomly distributed around the means. This 70% difference appeared to be caused by large differences when concentrations were <0.5 p.p.m. and when there were fewer than 30 measurements. At or above 1.0 p.p.m. and with 30 or more measurements the method performed well (within 60 and 37% of the observed means, respectively).

The 12 jobs with differences that exceeded 1 p.p.m. were examined for an explanation as to why the method performed poorly. Sixty percent were due to means that appeared to be outliers. For example, the Z department supervisor in plant 4 had a mean exposure of 3.15 p.p.m. ($n = 2$) in 1978. This value was twice as high as the mean of the next highest job (the senior operator, mean 1.69 p.p.m., $n = 40$) that same year and 15 times higher than the average exposure of the following year for the supervisor job (0.2 p.p.m., $n = 1$). Thus, the method does not appear to allow for particularly unusual situations. Although some jobs are likely to have exposures lower than expected, the major type of error for this model is likely to be biased towards overestimating the exposures.

*Homogeneous exposure group method*

Development of HEGs in this study assumed that descriptive information and observations of the workplace could be used to identify jobs that had similar exposures. If this assumption is incorrect, it is likely to result in random error. If, for example, a highly exposed job was inappropriately included in a low exposed HEG, the estimates for the other low exposed jobs in that HEG will be higher than the true exposure level. The estimate for the high exposed job, however, will be lower than the true exposure level.

Overall, the HEG method performed well. The estimates were within 60% of the means at concentrations equal to or exceeding 0.5 p.p.m. and the relative absolute differences indicated few large differences at these levels. Although the differences did not decrease as the concentrations increased, the relative absolute differences did. The method performed well regardless of the number of measurements when looking at simply the differences, but the relative absolute differences improved when there were 10 or more measurements. The six jobs for which the difference exceeded 1 p.p.m. were examined to determine if any patterns could be detected. All had unusual means. Half were also from one HEG. It appears, therefore, that the method performed poorly when the concentrations within an HEG were highly variable.

*Overall*

The estimates tended to be moderately correlated with the observed means ($r = 0.63–0.66$). The association between the concentration of the observed means and the differences was inconsistent across the methods, as was the correlation between the observed means and the relative absolute differences. The direction of all three sets of estimates tended to be higher than the observed means in the two lower exposure categories (<1 p.p.m.) and below 0.5 p.p.m. there were some large differences. Nevertheless, all of the methods resulted in generally good estimates (within 50–60% of the means) above 0.5 p.p.m.

The poor performance was expected at concentrations closer to the LOD. However, no pattern was found when the category was divided into values less than the LOD and greater than the LOD; the correlation coefficients for these categories were 0.14 versus 0.05 for the deterministic method, 0.20 versus 0.20 for the ratio method and –0.28 versus 0.32 for the HEG method. In addition, the mean differences were larger when the observed means were above the LOD but below 0.5 p.p.m. than when below the LOD. We have no explanation for this result. The results suggest, however, that the methods performed less well at lower levels. The magnitude of the relative differences is of less importance than the actual differences at these levels because the denominator used to calculate the relative differences is so small.

The magnitude of the differences was not statistically associated with the number of measurements for any of the methods. The analysis of the relative absolute difference, however, found that as the number of measurements increased, the differences decreased for the ratio and HEG methods. In addition, for all three methods the differences were within 30% of the means when the means were based on 30 or more measurements. This suggests that when 30 or more measurements are used to estimate exposure levels, the estimate may have higher validity than when the estimates are based on fewer measurements.

Few studies have been published that have evaluated occupational exposure estimation methods and these have not reported data that are easily comparable with the statistics evaluated in this paper. From a study of embalmers using a similar approach, we calculated the overall relative absolute difference from the paper (Hornung *et al.*, 1996) as 1.1. The bias of the deterministic and HEG methods was similar to that of the embalmers' study (1.0–1.1), while the bias of the ratio method was somewhat higher (1.6). In another study, industrial hygienists' estimates were compared with current styrene and methylene chloride exposure measurement results (Post *et al.*, 1991). The Spearman correlation coefficients between the three raters' rankings and the ranking of the jobs by measured exposure concentration were between 0.67 and 0.73 for methylene chloride and between 0.12 and 0.29 for styrene. In comparison, this study of acrylonitrile exposures found correlations of about 0.65 between the measurement data and the estimates of all three methods. Considering that the methylene chloride and styrene evaluations were based on current jobs by raters on site, whereas this evaluation was based on exposures occurring up to 16 yr prior to the estimation effort, the acrylonitrile results compare quite favorably. More recently a deterministic approach incorporating evaluations of emissions (characteristic of the substance, handling and duration of emissions), use of protective equipment and ventilation characteristics was used to estimate exposure for five chemicals (Cherrie and Schneider, 1999). The correlations between the estimates and the measurement results ranged from 0 to 0.93 (mean 0.46, median 0.50).

This evaluation used measurement data as a surrogate for the true exposures. In many cases, few data were available to represent these exposures, and we found that the number of measurements affected to some degree the relative performance of all three of the estimation methods. It is possible that the methods used here actually did better in predicting the true exposures than indicated by the measured means or, at a minimum, that the methods may have smoothed the exposure trends across jobs or over time. Support for this reasoning is that in several of the comparisons where the methods performed particularly poorly, the measurement results were substantially different from other years for the same job or from other jobs to which the exposures were similar in other years. These unexpected measurement results may reflect actual annual exposures, but they may also be the result of unrepresentative monitoring, which can have greater impact on a mean when the number of measurements is small. This finding is not unique, however, to the estimation methods described here; rather it is typical of all models. Although more work needs to be done to verify this finding, caution is recommended when using measurement data to estimate highly variable jobs, particularly if few data are available. Weighting of the measurements based on frequency of unusual occurrences can be done to reduce this possible source of bias (Stewart *et al.*, 1996b). Caution is also advised concerning the precision of the estimates reported. The means in the tables were presented to the level of precision reported by the laboratory, but cannot be assumed to be the precise differences between the methods and the measurement data. Rather, they represent relative and approximate values. The use of such data is likely to have minimal effect on the estimated differences.

Few epidemiological studies have had as many measurements and as many observations to evaluate the assessment methods as this one. In spite of the large numbers, there was only a limited number of comparisons that could be examined after time periods were developed that represented periods of stable exposures and after means had been calculated for individual jobs within those time periods. Moreover, because of the lack of independence among the methods, the power to distinguish among the methods was limited. It is also not known whether these results are generalizable to other exposures or other industries.

The credibility of an epidemiological study rests strongly on the credibility of the exposure estimates. Lack of a validity evaluation reduces the usefulness of the study and opens the investigators to criticism. It is recommended that every epidemiological study that investigates disease risks from occupational hazards evaluate the validity of the exposure assessments, wherever possible. Such an evaluation should be done prior to the development of the estimates to allow the investigators to modify the estimation methods if poor estimates are developed. The approach can be similar to that taken here, i.e. removing a subset of the measurement data, estimating the exposures and comparing the estimates to the measurement results. If this exercise is done prior to developing estimates for the study, all measurement data are available for the estimation procedure. The degree of misclassification found can then be applied to the expected exposure–response relationships to determine if the interpretation of the data is adversely affected.

## APPENDIX

Table A1 shows the observed results from the estimation methods used.

Table A1. The observed means and estimates from the three estimation methods

| Observed mean (p.p.m.) | No. of observed measurements | Determined estimate (p.p.m.) | No. of measurements for determined estimate | Ratio estimate (p.p.m.) | No. of measurements for ratio estimates | HEG estimate (p.p.m.) | No. of measurements for HEG estimates | Job title | Process (plant no.) | Time period |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.1 | 22 | 3.43 | 166 | 0.76 | 69 | 1.07 | 48 | Polymer operator I | Fiber (1) | 1978–1978 |
| 1.3 | 28 | 1.61 | 166 | 1.05 | 72 | 4.20 | 48 | Polymer operator I | Fiber (1) | 1979–1979 |
| 0.41 | 24 | 1.59 | 166 | 0.25 | 79 | 0.70 | 54 | Polymer operator I | Fiber (1) | 1980–1980 |
| 1.87 | 10 | 0.99 | 166 | 0.57 | 32 | 0.93 | 30 | Polymer operator I | Fiber (1) | 1981–1981 |
| 0.76 | 27 | 2.97 | 139 | 1.10 | 64 | 1.28 | 43 | Polymer operator II | Fiber (1) | 1978–1978 |
| 1.05 | 24 | 1.33 | 139 | 1.30 | 76 | 4.10 | 52 | Polymer operator II | Fiber (1) | 1979–1979 |
| 0.25 | 29 | 1.32 | 139 | 0.41 | 74 | 0.82 | 49 | Polymer operator II | Fiber (1) | 1980–1980 |
| 0.57 | 14 | 0.82 | 139 | 1.87 | 28 | 1.48 | 30 | Polymer operator II | Fiber (1) | 1981–1981 |
| 1.48 | 21 | 3.52 | 108 | 1.58 | 70 | 0.91 | 49 | Assistant polymer operator | Fiber (1) | 1978–1978 |
| 7.36 | 24 | 1.82 | 108 | 7.54 | 76 | 1.18 | 52 | Assistant polymer operator | Fiber (1) | 1979–1979 |
| 1.22 | 25 | 1.80 | 108 | 1.37 | 78 | 0.32 | 53 | Assistant polymer operator | Fiber (1) | 1980–1980 |
| 1.5 | 9 | 1.12 | 108 | 2.69 | 33 | 1.07 | 31 | Assistant polymer operator | Fiber (1) | 1981–1981 |
| 3.19 | 21 | 1.77 | 78 | 2.09 | 39 | 2.57 | 54 | B polymer operator | Fiber (4) | 1977–1977 |
| 1.8 | 58 | 1.64 | 78 | 2.00 | 72 | 2.02 | 114 | B polymer operator | Fiber (4) | 1978–1978 |
| 1.38 | 61 | 1.62 | 78 | 1.51 | 76 | 1.90 | 128 | B polymer operator | Fiber (4) | 1979–1980 |
| 2.83 | 30 | 2.26 | 83 | 4.31 | 30 | 2.68 | 45 | B salt handler | Fiber (4) | 1977–1977 |
| 2.13 | 56 | 2.10 | 83 | 3.05 | 74 | 1.86 | 116 | B salt handler | Fiber (4) | 1978–1978 |
| 1.86 | 69 | 2.07 | 83 | 1.94 | 77 | 1.67 | 120 | B salt handler | Fiber (4) | 1979–1980 |
| 1.92 | 58 | 2.38 | 79 | 1.26 | 72 | 1.96 | 114 | B salt purifier | Fiber (4) | 1978–1978 |
| 1.96 | 59 | 2.34 | 79 | 1.73 | 114 | 1.63 | 130 | B salt purifier | Fiber (4) | 1979–1980 |
| 1.56 | 10 | 4.70 | 87 | 0.42 | 84 | 1.31 | 45 | B assistant head operator | Fiber (4) | 1978–1978 |
| 1.69 | 61 | 4.12 | 87 | 0.90 | 42 | 1.81 | 142 | B assistant head operator | Fiber (4) | 1979–1980 |
| 0.33 | 4 | 3.09 | 47 | 1.19 | 90 | 1.43 | 51 | B head operator | Fiber (4) | 1978–1978 |
| 0.84 | 8 | 2.71 | 47 | 1.58 | 95 | 1.81 | 195 | B head | Fiber (4) | 1979–1980 |

Table A1. *Continued*

| Observed mean (p.p.m.) | No. of observed measurements | Determined estimate (p.p.m.) | No. of measurements for determined estimate | Ratio estimate (p.p.m.) | No. of measurements for ratio estimates | HEG estimate (p.p.m.) | No. of measurements for HEG estimates | Job title | Process (plant no.) | Time period |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.24 | 8 | 0.26 | 3 | 0.12 | 8 | 0.15 | 9 | B dry tow handler | Fiber (4) | 1978–1978 |
| 0.16 | 4 | 1.41 | 1 | 0.22 | 11 | 0.20 | 13 | B baler | Fiber (4) | 1978–1978 |
| 0.83 | 4 | 1.37 | 1 | 7.35 | 10 | 3.13 | 7 | B supervisor | Fiber (4) | 1978–1978 |
| 0.82 | 6 | 0.21 | 2 | 1.03 | 7 | 0.84 | 15 | B assistant superintendent | Fiber (4) | 1978–1978 |
| 1.13 | 4 | 0.19 | 1 | 0.47 | 9 | 0.76 | 17 | B superintendent | Fiber (4) | 1978–1978 |
| 0.98 | 6 | 0.50 | 88 | 0.94 | 124 | 1.11 | 18 | Z polymer operator | Fiber (4) | 1978–1978 |
| 0.49 | 16 | 0.53 | 88 | 0.44 | 121 | 0.62 | 18 | Z polymer operator | Fiber (4) | 1979–1979 |
| 0.54 | 27 | 0.52 | 88 | 0.52 | 146 | 0.71 | 56 | Z polymer operator | Fiber (4) | 1980–1981 |
| 1.39 | 10 | 0.81 | 85 | 0.99 | 120 | 0.85 | 14 | Z salt handler | Fiber (4) | 1978–1978 |
| 0.62 | 17 | 0.85 | 85 | 0.61 | 129 | 0.50 | 17 | Z salt handler | Fiber (4) | 1979–1979 |
| 0.74 | 30 | 0.84 | 85 | 0.67 | 154 | 0.61 | 53 | Z salt handler | Fiber (4) | 1980–1981 |
| 0.75 | 8 | 0.88 | 82 | 1.15 | 122 | 1.24 | 16 | Z salt purifier | Fiber (4) | 1978–1978 |
| 0.6 | 1 | 0.92 | 82 | 0.68 | 136 | 0.56 | 33 | Z salt purifier | Fiber (4) | 1979–1979 |
| 0.68 | 26 | 0.91 | 82 | 0.79 | 147 | 0.65 | 57 | Z salt purifier | Fiber (4) | 1980–1981 |
| 1.69 | 40 | 0.84 | 47 | 6.19 | 54 | 1.39 | 61 | Z senior operator | Fiber (4) | 1978–1978 |
| 0.53 | 19 | 0.85 | 47 | 0.97 | 84 | 0.67 | 66 | Z senior operator | Fiber (4) | 1979–1979 |
| 0.61 | 30 | 0.76 | 47 | 1.05 | 96 | 0.78 | 93 | Z senior operator | Fiber (4) | 1980–1981 |
| 0.48 | 15 | 0.78 | 71 | 0.26 | 88 | 0.68 | 70 | Z head operator | Fiber (4) | 1979–1979 |
| 0.6 | 27 | 0.70 | 71 | 0.35 | 99 | 0.78 | 96 | Z head operator | Fiber (4) | 1980–1981 |
| 0.15 | 2 | 0.15 | 25 | 0.30 | 14 | 0.21 | 7 | Z dry tow handler | Fiber (4) | 1978–1978 |
| 0.3 | 1 | 0.15 | 25 | 0.10 | 6 | 0.17 | 3 | Z dry tow handler | Fiber (4) | 1979–1979 |
| 0.26 | 4 | 0.15 | 2 | 0.18 | 11 | 0.15 | 5 | Z cutter | Fiber (4) | 1978–1978 |
| 0.2 | 2 | 0.20 | 1 | 0.20 | 13 | 0.20 | 7 | Z baler | Fiber (4) | 1978–1978 |
| 3.15 | 2 | 0.45 | 3 | 0.35 | 12 | 0.49 | 7 | Z area supervisor | Fiber (4) | 1978–1978 |
| 0.2 | 1 | 0.46 | 1 | 0.02 | 4 | 0.07 | 2 | Z area supervisor | Fiber (4) | 1979–1979 |
| 0.1 | 1 | 0.43 | 1 | 0.05 | 4 | 0.82 | 6 | Z supervisor | Fiber (4) | 1980–1981 |
| 0.35 | 2 | 0.45 | 3 | 1.63 | 12 | 0.12 | 4 | Z assistant superintendent | Fiber (4) | 1978–1978 |
| 0.02 | 1 | 0.06 | 3 | 0.2 | 4 | 0.31 | 6 | Z assistant superintendent | Fiber (4) | 1979–1979 |
| 0.2 | 1 | 0.06 | 3 | 0.48 | 12 | 0.19 | 5 | Z superintendent | Fiber (4) | 1978–1978 |

Table A1. *Continued*

| Observed mean (p.p.m.) | No. of observed measurements | Determined estimate (p.p.m.) | No. of measurements for determined estimate | Ratio estimate (p.p.m.) | No. of measurements for ratio estimates | HEG estimate (p.p.m.) | No. of measurements for HEG estimates | Job title | Process (plant no.) | Time period |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.4 | 1 | 0.43 | 15 | 9.66 | 10 | 1.38 | 5 | Reactor, chemical operator | Production (3) | 1978–1978 |
| 0.15 | 4 | 0.45 | 15 | 0.38 | 29 | 0.74 | 8 | Reactor, chemical operator | Production (3) | 1980–1980 |
| 0.2 | 2 | 0.39 | 15 | 0.24 | 29 | 0.61 | 10 | Reactor, general operator | Production (3) | 1980–1980 |
| 1.1 | 3 | 1.25 | 23 | 25.81 | 10 | 1.33 | 3 | Purification, chemical operator | Production (3) | 1978–1978 |
| 1.25 | 2 | 1.26 | 23 | 0.82 | 44 | 0.40 | 10 | Purification, chemical operator | Production (3) | 1980–1980 |

## REFERENCES


Cherrie JW, Schneider T. (1999) Validation of a new method for structured subjective assessment of past concentrations. Ann Occup Hyg; 43: 235–45.

Efron B, Gong G. (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. Am Statistician; 37: 36–48.

Eisen EA, Smith TJ, Wegman DH, Louis TA, Froines J. (1984) Estimation of long term dust exposures in the Vermont granite sheds. Am Ind Hyg Assoc J; 45: 89–94.

Esmen N. (1979) Retrospective industrial hygiene surveys. Am Ind Hyg Assoc J; 40: 58–65.

Hornung RW, Reed LD. (1990) Estimation of average concentration in the presence of nondetectable values. Appl Occup Environ Hyg; 5: 46–51.

Hornung RW, Herrick RF, Stewart PA, Utterback DF, Feigley CE, Wall DK, Douthit DE, Hayes RB. (1996) An experimental design approach to retrospective exposure assessment. Am Ind Hyg Assoc J; 57: 251–6.

Post W, Kromhout H, Heederik D, Noy D, Duijzentkunst RS. (1991) Semiquantitative estimates of exposure to methylene chloride and styrene: the influence of quantitative exposure data. Appl Occup Environ Hyg; 6: 197–204.

Seixas NS, Robins TG, Mouton LH. (1988) The use of geometric and arithmetic mean exposures in occupational epidemiology. Am J Ind Med; 14: 465–77.

Snedecor GW, Cochran WG. (1973) Statistical methods. Ames, IA: Iowa State University Press, p. 128.

Stewart PA, Lees PS, Francis M. (1996a) Quantification of historical exposures in occupational cohort studies. Scand J Work Environ Health; 22: 405–14.

Stewart PA, Zey JN, Hornung R, Herrick RF, Dosemeci M, Zaebst D, Pottern LM. (1996b) Exposure assessment for a study of workers exposed to acrylonitrile. III. Evaluation of exposure assessment methods. Appl Ind Hyg; 11: 1312–21.

Zey JN, Stewart PA, Hornung R, Herrick R, Mueller CA, McCammon C, Zaebst D, Pottern LM, Dosemeci M, Bloom T. (2002) Evaluation of concurrent measurements of acrylonitrile using different sampling techniques. Appl Occup Environ Hyg; 17: 88–95.